DOCUMENT RESUME

ED 049 890 RE 003 422

AUTHOR Pellegrine, R. J.
TITLE The Diagnostic Reading Test, Survey Section, Form E: A Reliability Study.
PUB DATE Dec 70
NOTE 16p.; Paper presented at the National Reading Conference, St. Petersburg, Fla., Dec. 3-5, 1970

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *College Freshmen, Disadvantaged Youth, *Reading Centers, Reading Diagnosis, *Reading Research, *Reading Tests, *Test Reliability

ABSTRACT

The Diagnostic Reading Tests were designed to assess the reading skills of college students enrolled in reading centers. To assess the reliability of the Diagnostic Reading Tests, Survey Section, Form E (DRTE), a study was conducted with university freshmen as subjects. The DRTE was administered to 31 students in an Educational Opportunity Program (EOP), to 122 students admitted to the Eastman School of Music of the University (ESM), and to 229 students who were regularly admitted to attend the Summer Orientation Program (SOP) session. The DRTE was successful in differentiating the three groups--the SOP group had the highest mean scores, as was expected; the EOP group had the lowest, with the ESM group somewhere in the middle. Reliability estimates of the DRTE using the Kuder-Richardson Formulas 20 and 21 were found to be satisfactory and did not differ from those published by the test authors. The fact that 16 of the test items did not contribute significantly to the test score seemed to indicate that a revision of the test is needed. Tables and references are included. (DH)

ED049890

# The Diagnostic Reading Test, Survey Section, Form E: A Reliability Study.

R. J. Pellegrine
Rochester University

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

In reading, as in any other area, we are interested in obtaining the most accurate estimates of an individual's abilities. With the ever increasing numbers of students attending colleges and junior colleges, an ever increasing number of reading centers are being established to meet the academic needs of students. Programs designed to provide individual reading assistance demand a continuously monitored accurate estimate of reading skills. The Diagnostic Reading Tests were designed for that purpose. The present study was undertaken to assess the reliability estimate of the DRT Survey Section, Form E (DRTE) on a sample of a university Freshman population.

The assessment of reading strengths and weaknesses depends heavily upon measuring instruments that are reliable. A reliable instrument will measure behavior consistently over a series of measurements. Reliable instruments supply credible information about an individual, credible in that the results are consistently accurate within acceptable limits. This credibility is achieved through the control of error variance. We seek and need instruments in which the total variance is not heavily confounded with sources of error variance, especially error variance

RE003 422

attributable to the instrument inself. A reliable instrument then is one in which the estimate of true variance is accurate.

Methods of estimating reliability are too well known to be described in depth here. It is incumbent upon test authors to report how they made their estimate of the reliability coefficient, since the methods differ such that "no close estimate can be made of the results of one method, knowing the estimate obtained by another method." (4, p. 151) Test-retest usually results in reliability coefficients that are elevated because of both practice effect and remembering materials on the second administration. Increasing the interval between tests does not appear to eliminate the memory factor. This contamination seriously limits the utilization of this method.

Split half methods of reliability seem most popular. The investigator arbitrarily divides the test into two parts and computes correlation coefficients for the two sets of scores. The Spearman-Brown formula is then employed to estimate the reliability coefficient for the whole test. The confusion here is that there are many different ways of dividing a test in half. Each division will supply a different and valid reliability coefficient. Brownell (1) demonstrated that large fluctuations appear in the coefficients obtained from the different ways of splitting the same instrument.

Equivalent forms appear to provide about the most accurate estimate of reliability. The question here is equivalent. Estimates obtained by so-called equivalent forms are in fact as variable as those obtained by split halves. We may have n equivalent forms and obtain a variable number of valid coefficients of reliability.

More recently, it has become fashionable to employ Kuder-Richardson formulae in estimating the reliability coefficient of measuring instruments. Two such formulae are KR 20 and KR 21 (4). Formula 20 may be employed when the items of a test appear to be of equal difficulty. Formula 21 is used when the item difficulty appears to vary greatly.

## METHOD

### Subjects

The organizational structure of the University of Rochester offered a unique opportunity to test three distinct groups of students with the DRTE during the summer of 1969. At that time, the University admitted approximately fifty students who would normally not meet admission requirements based on academic achievement and aptitude test scores (CEEB's). Those individuals were admitted into the Educational Opportunity Program (EOP). Students selected for the EOP were minority group inner city youths who possessed the potential for success in academic pursuits. During the summer of 1969, these EOP students were introduced to the university and exposed to programs designed to equip them with the skills and experiences requisite for a successful academic career.

A second unique group was composed of students admitted to the Eastman School of Music of the University. These students were and are usually matriculated on the basis of musical potential first, academic ability second.

The third group of students were the regularly admitted incoming freshmen who elected to attend a particular Summer Orientation Program (SOP) Session.

Procedure and Results

The DRTE was administered to 31 EOP, 122 ESM and 229 SOP students of both sexes admitted to the University of Rochester during the summer of 1969. Test answer sheets were scored according to manual instructions. Means, standard deviations and values of *t* were computed for each group and are presented in Table I.

---

Table I about here

---

From Table I it is obvious that the three groups differed significantly. The SOP group earned the highest mean score and the lowest standard deviation. They are a high ability group. Examination of test answer sheets revealed many SOP students achieved perfect comprehension scores on part 1b, reading comprehension.

The ESM students earned a mean score significantly less than that of the SOP group. The standard deviation for the ESM group was also greater, indicating a wider range of scores. They are a median ability group.

Finally, the EOP group earned the lowest mean score and the highest standard deviation, indicating a low ability group, widely dispersed.

Values of *t* were significant at or beyond the .001 level between the three groups, indicating significantly different reading ability groups.

Inspection of published normative data indicated the three groups, combined, represent a close approximation of usual entering freshmen. With this in mind, the three groups were combined and point biserial coefficients of correlation were computed for each test item with total

score. The correlations of 16 test items did not attain the .05 level of significance. Those items did not contribute significantly to the total score. The 16 items and corresponding point biserial coefficients of correlation are presented in Table II.

---

Table II about here

---

Table II includes one item from each comprehension section and 14 items from the vocabulary section. The 16 noncontributory items might well be dropped from the instrument. To test this, test answer sheets were rescored deleting the 16 responses. A Pearson Product Moment Coefficient of Correlation was computed between the two sets of scores and was of the order of .99. A linear transformation had been accomplished. It was concluded that the two test forms were equivalent.

The Committee of Diagnositc Reading Tests (2) reports estimates of reliabilities for the DRTE as follows:

| Part | Method | Median r |
|---|---|---|
| 1a | Test-retest | .80 |
| 1b | KR 21 | .74 |
| 2 | KR 21 | .89 |
| 3 | KR 21 | .83 |
| Total | KR 21 | .91 |

Confusion exists in the author's mind as to the population tested. The Committee (2, p. 9) reports a sample drawn from grades 7-12, with 100 individuals in each group. The same manual (Committee, 1952, p. 19) reports estimates that appear to have been computed from test results of a different group, a population including college freshmen. The reported reliabilities are for forms A and B only. They suggest "These reliabilities are typical of other forms of the test." (2, p. 19). I

have been unable to locate any information concerning the reliability of the DRTE. I felt that the above statement by the Committee was insufficient evidence as to the reliability of the E form of the test.

In the present study, two methods of estimating reliability coefficients were employed. Since the methods differ, by definition the obtained estimates should differ. The two methods used were the Kuder-Richardson formulae 20 and 21 (KR 20, KR 21). Cronbach (3) suggests the Kuder-Richardson Formula 21 is a less accurate procedure but simpler to compute. It is a formula that "...can be used by any tester to get quick estimates of the coefficient of equivalence in his group." (3, p. 141). KR 21 employs the mean, the number of items, and the standard deviation. It generally will generate a coefficient that is very nearly the same as the KR 20 but will sometimes generate a much lower one. The KR 20 formula is considered to generate a coefficient that will be "a good approximation of an equivalent-form correlation." (3, p. 141). KR 20 employs the proportion of individuals passing each item and the standard deviation.

Reliability coefficients were computed for the full length test and are presented in Table III.

---

Table III about here

---

Coefficients derived by the two formulae do not differ significantly. It is apparent that the reliability of the DRTE is satisfactory. The most reliable section of the test continues to be Vocabulary. Utilizing the KR 20 formula, reliabilities are elevated, but only slightly. It was concluded that the DRTE is a reliable instrument.

As a further part of the present study, the same reliability formulae were applied to the 84-item form of the test. The resultant coefficients are presented in Table IV.

---

Table IV about here

---

The short form of the instrument is as reliable as the long form. This seems to coincide with the previous finding (in this study) of a correlation between the two forms. The 84-item form of the DRTE may be a reliable instrument.

Pearson Product Moment Coefficients of Correlation were computed to determine the relationship between part and total scores and are presented in Table V.

---

Table V about here

---

From Table V all parts of the DRTE are interrelated minimally. The degree of intercorrelation indicates the separate scores are measures of separate variables. The high correlation between parts 1b and 3 was accounted for by the fact that score 3 includes score 1b. The relationship between the vocabulary scores, score 2 and the total score indicates that perhaps the two are measuring the same variable. It must be remembered that the vocabulary section of the DRTE contains three times as many items as part 1b, and one third more items than part 3. This loading is represented in the mentioned correlation.

All of the coefficients of correlation calculated here do not differ from those published by the test authors (2, p. 13).

## DISCUSSION

The freshman population of the University of Rochester is composed of three distinct academic ability groups. Regularly admitted freshman students come from highly verbal families and are well prepared for academic ventures. They place relatively high on College Boards and constitute a select sample of bright verbal young men and women. Students admitted to the Eastman School of Music of the University are less well prepared academically. Their strength lies primarily in musical potential. They constitute a middle group as far as academic ability is concerned. The lowest or poorest prepared group are the EOP freshmen. Most of them came from inner city schools. College Board scores for this group are relatively low. The DRTE seems to have been successful as an instrument to differentiate these groups. As might be expected, the SOP group had the highest DRTE mean scores, the EOP group had the lowest, with the ESM group somewhere in the middle. Overlapping seems evident but it was apparent from the distribution of scores that few of the EOP group attained mean score level for the combined group. Relatively few of the SOP group fell at or near the combined group mean. The DRTE might be used as an instrument to identify students of promise. It must be recalled that previous studies attempting to relate reading test scores with academic success have been less than successful. The variable of motivation has been elusive and evades control. Until such studies are designed, success in this line of research will be extremely limited. The present findings indicate that those students who come from verbal families and who attended the "better" secondary schools score higher on the DRTE.

The disclosure that 16 of the 100 test items did not contribute

significantly to total score indicates that revision of the instrument is or may be needed. Verification of the above findings with another sample of college freshmen would be additional evidence indicating revision was needed. The present author is in the process of collecting data for such a study.

The reduction of the DRTE by 16 items raises some administrative problems. First, only one item from each of the two comprehension sections would be deleted. Attempting to reduce the test administration time by an appropriate amount would be difficult. The greatest saving in administration time is with the vocabulary section. Deletion of the 14 non-contributory items results in a 23 percent reduction of the instruments and a 23 percent reduction in test time. Test scoring time would also be reduced. This time would be saved without significant loss of data. Obviously some students will be penalized by the test item reduction. But the relationship between the instruments appears to be large enough that this would penalize a very small number of students. It seems essential that careful attention should be focused on replicating the findings and verifying the results prior to such revision.

The present study relied on a statistical manipulation of test answer sheets in arriving at the 84-item test. Test answer sheets were rescored, deleting the 16 items. When the correlation was computed, the author felt that the test should be duplicated, deleting the 16 items and administered to another group of students. He sought the cooperation of the Committee on Diagnostic Reading Tests for such permission. The Chairman of the Committee notified him that the policy of the Committee precluded this and permission was not granted. The chairman refused to grant permission and cooperate with this research.

## SUMMARY

The purpose of the present study was to evaluate the reliability of the DRTE on a university freshman sample. It was determined that the instrument was reliable. It was further determined that the instrument might be shortened and still maintain reliability. Verification of the results depends upon the cooperation of the Committee..."The plan of the Committee is to provide for continuous research on the tests and for revision of them as needed." (5) That time seems to have arrived.

1. Brownell, W. A. On the accuracy with which reliability may be measured by correlating test halves. Journal of Experimental Education, 1933, 1, 204-215.

2. Committee on Diagnostic Reading Tests: Diagnostic Reading Tests: A History of their Construction and Validation: New York, The Committee, 1952.

3. Cronbach, L. J. Essentials of Psychological Testing, New York: Harper and Row, 1960.

4. Kuder, G. F., and M. W. Richardson. The Theory of the estimation of test Reliability. Psychometrics, 1937, 2 (3), 151-160.

5. Triggs, F. O. Description of the Purposes and Functions of The Diagnostic Reading Tests. Educational and Psychological Measurement, 1948, 8, 3-14.

Table I

Mean Standard Deviation and values of t of DRT total scores for three groups.

| Group | $\bar{x}$ | SD | t | df |
|---|---|---|---|---|
| ESM<br>SOP | 78.24<br>86.06 | 10.10<br>7.15 | 8.383* | 349 |
| ESM<br>EOP | 78.24<br>63.93 | 10.10<br>11.16 | 6.842* | 151 |
| EOP<br>SOP | 63.93<br>86.06 | 11.16<br>7.15 | 14.882* | 258 |

* $p < .001$

Table II

Point Biserial Coefficients of correlation for 16 test items.

| Item # | $r_{pb}$ |
|---|---|
| 7 | 1.559 |
| 21 | 1.095 |
| 24 | 0.993 |
| 27 | 0.952 |
| 30 | 0.187 |
| 31 | 1.962 |
| 32 | 1.010 |
| 33 | 0.174 |
| 36 | 1.819 |
| 37 | 0.300 |
| 39 | 1.448 |
| 43 | 1.610 |
| 45 | 1.498 |
| 47 | 1.029 |
| 87 | 0.253 |
| 91 | 0.490 |

Table III

Reliability Coefficients Derived by 2 Formulae on 100 Item DRTE

| Part | Formula KR 21 | KR 20 | N |
|---|---|---|---|
| 1b | .66 | .70 | 382 |
| 2 | .93 | .94 | |
| 3 | .63 | .70 | |
| Total | .90 | .92 | |

Table IV

Reliability Coefficients Derived by 2 Formulae on 84 Item DRTE

| Part | Formula KR 21 | KR 20 | N |
|---|---|---|---|
| 1b | .66 | .69 | 382 |
| 2 | .95 | .96 | |
| 3 | .65 | .72 | |
| Total | .91 | .93 | |

Table V

Pearson Product moment coefficients of correlation between parts and total scores in DRTE.

| Part | 1b | 2 | 3 | n |
|---|---|---|---|---|
| 1b | | | | 382 |
| 2 | 0.421* | | | |
| 3 | 0.887* | 0.538* | | |
| Total | 0.663* | 0.940* | 0.793* | |

*Sig> .01

Table VI

MEANS, STANDARD DEVIATION AND STANDARD ERROR OF MEASUREMENT--DRTE
KR 20 and KR 21 Combined Group

| Part | Mean | SD | Sem(KR20) | Sem(KR21) |
|---|---|---|---|---|
| 1b | 15.43 | 2.98 | 1.63 | 1.72 |
| 2 | 49.50 | 7.72 | 1.93 | 2.03 |
| 3 | 32.27 | 4.31 | 2.23 | 2.62 |
| Total | 81.77 | 10.68 | 3.03 | 3.38 |